microPublication BIOLOGY

9/14/2025 - *Open Access*

Gene model for the ortholog of *dock* in *Drosophila ananassae*

Anne E. Backlund[1], Tamica D'Souza[2], Jaskamaldip Kaur[3], Anya Goodman[2], James J. Youngblom[3], Chinmay P. Rele[1], Laura K Reed[1§]

[1]The University of Alabama, Tuscaloosa, AL USA

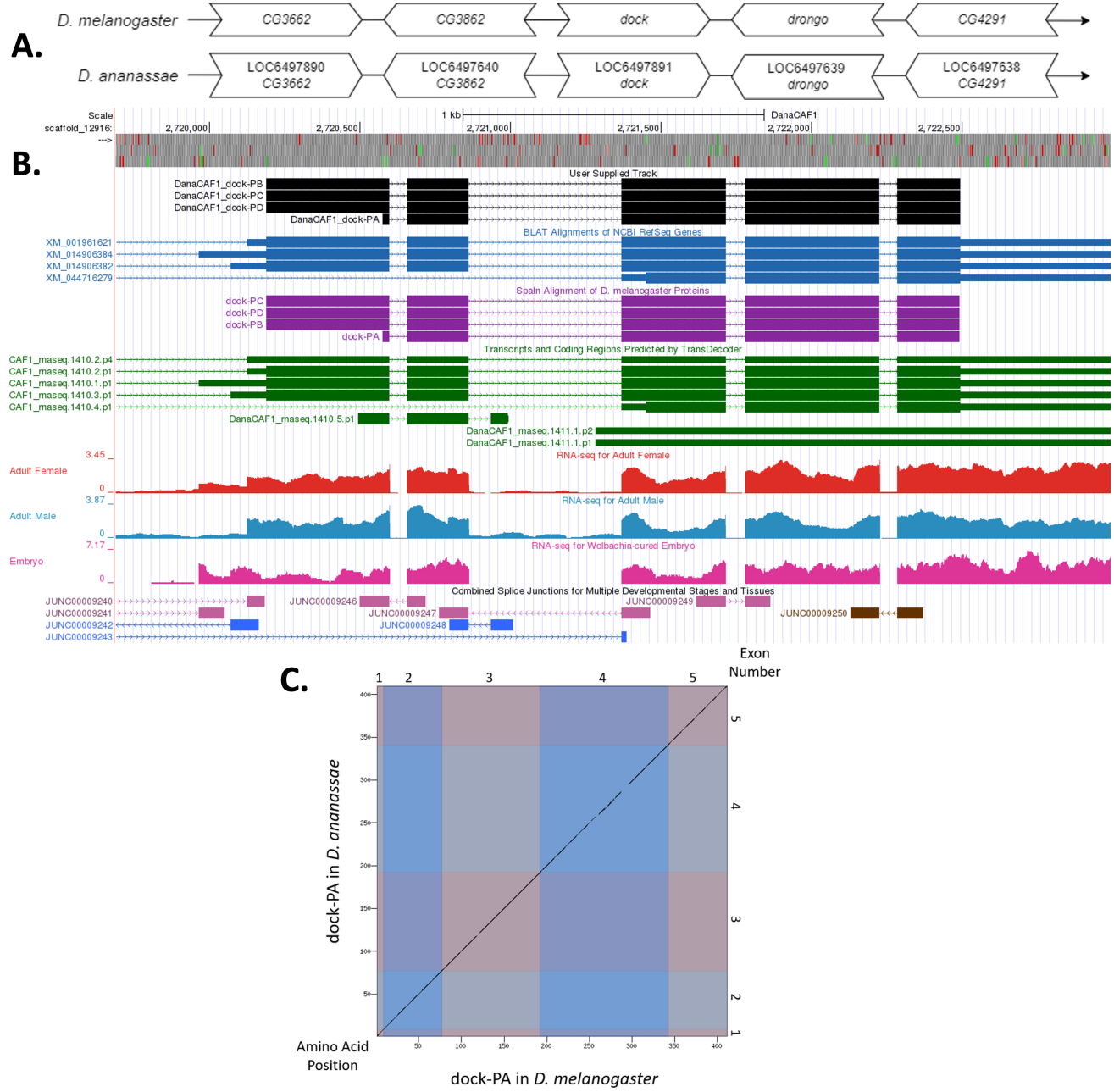
[2]California Polytechnic State University, San Luis Obispo, CA USA

[3]California State University Stanislaus, Turlock, CA USA

[§]To whom correspondence should be addressed: lreed1@ua.edu

Abstract

Gene model for the ortholog of *dreadlocks* (*dock*) in the May 2011 (Agencourt dana_caf1/DanaCAF1) Genome Assembly (GenBank Accession: GCA_000005115.1) of *Drosophila ananassae*. This ortholog was characterized as part of a developing dataset to study the evolution of the Insulin/insulin-like growth factor signaling pathway (IIS) across the genus *Drosophila* using the Genomics Education Partnership gene annotation protocol for Course-based Undergraduate Research Experiences.



9/14/2025 - *Open Access*

Figure 1. Genomic neighborhood and gene model for *dock* in *Drosophila ananassae*:

(A) Synteny comparison of the genomic neighborhoods for *dock* in *Drosophila melanogaster* and *D. ananassae*. Thin underlying arrows indicate the DNA strand within which the gene–*dock*–is located in *D. melanogaster* (top) and *D. ananassae* (bottom). The thin arrows pointing to the right indicate that *dock* is on the positive (+) strand in *D. ananassae* and *D. melanogaster*. The wide gene arrows pointing in the same direction as *dock* are on the same strand relative to the thin underlying arrows, while wide gene arrows pointing in the opposite direction of *dock* are on the opposite strand relative to the thin underlying arrows. White gene arrows in *D. ananassae* indicate orthology to the corresponding gene in *D. melanogaster*. Gene symbols given in the *D. ananassae* gene arrows indicate the orthologous gene in *D. melanogaster*, while the locus identifiers are specific to *D. ananassae*. **(B) Gene Model in GEP UCSC Track Data Hub** (Raney et al., 2014). The coding-regions of *dock* in *D. ananassae* are displayed in the User Supplied Track (black); coding CDSs are depicted by thick rectangles and introns by thin lines with arrows indicating the direction of transcription. Subsequent evidence tracks include BLAT Alignments of NCBI RefSeq Genes (dark blue, alignment of Ref-Seq genes for *D. ananassae*), Spaln of *D. melanogaster* Proteins (purple, alignment of Ref-Seq proteins from *D. melanogaster*), Transcripts and Coding Regions Predicted by TransDecoder (dark green), RNA-Seq from Adult Females, Adult Males, and Wolbachia-cured Embryos (red, light blue, and pink, respectively); alignment of Illumina RNA-Seq reads from *D. ananassae*), and Splice Junctions Predicted by regtools using *D. ananassae* RNA-Seq (SRP006203, SRP007906, PRJNA257286, PRJNA388952). Splice junctions shown have a minimum read-depth of 10 with 10-49, 100-499, 500-999 supporting reads in blue, pink, and brown, respectively. **(C) Dot Plot of dock-PA in *D. melanogaster* (*x*-axis) vs. the orthologous peptide in *D. ananassae* (*y*-axis).** Amino acid number is indicated along the left and bottom; CDS (exon) number is indicated along the top and right, and CDSs are also highlighted with alternating colors. Line breaks in the dot plot indicate mismatching amino acids at the specified location between species.

Description

This article reports a predicted gene model generated by undergraduate work using a structured gene model annotation protocol defined by the Genomics Education Partnership (GEP; thegep.org) for Course-based Undergraduate Research Experience (CURE). The following information in this box may be repeated in other articles submitted by participants using the same GEP CURE protocol for annotating Drosophila species orthologs of Drosophila melanogaster genes in the insulin signaling pathway.

"In this GEP CURE protocol students use web-based tools to manually annotate genes in non-model *Drosophila* species based on orthology to genes in the well-annotated model organism fruitfly *Drosophila melanogaster*. The GEP uses web-based tools to allow undergraduates to participate in course-based research by generating manual annotations of genes in non-model species (Rele et al., 2023). Computational-based gene predictions in any organism are often improved by careful manual annotation and curation, allowing for more accurate analyses of gene and genome evolution (Mudge and Harrow 2016; Tello-Ruiz et al., 2019). These models of orthologous genes across species, such as the one presented here, then provide a reliable basis for further evolutionary genomic analyses when made available to the scientific community." (Myers et al., 2024).

"The particular gene ortholog described here was characterized as part of a developing dataset to study the evolution of the Insulin/insulin-like growth factor signaling pathway (IIS) across the genus *Drosophila*. The Insulin/insulin-like growth factor signaling pathway (IIS) is a highly conserved signaling pathway in animals and is central to mediating organismal responses to nutrients (Hietakangas and Cohen 2009; Grewal 2009)." (Myers et al., 2024).

"The product of gene *dreadlocks* (*dock*, FBgn0010583) is involved in several cellular functions, including axon guidance (Garrity et al., 1996; Hing et al., 1999; Schmucker et al., 2000; Stevens and Jacobs 2002; Weng et al., 2011), myoblast fusion during muscle fiber formation (Kaipa et al., 2013), regulation of intercellular bridges in germline cells during gametogenesis (Stark et al., 2021), and negative regulation of insulin receptor signaling pathway (Wu et al., 2011; Willoughby et al., 2017). A *dock* transcript was first isolated and sequenced in *Drosophila melanogaster* in a screen for P-element insertions leading to R cell projection defects (Garrity et al., 1996). *dock* encodes a protein that contains three N-terminal SH3 domains and one C-terminal SH2 domain known to bind to specific motifs and serve as binding adapters (Garrity et al., 1996). Its regulation of photoreceptor axon guidance in *Drosophila* occurs through its interaction with InR (Rao et al., 1998; Song et al., 2003; Rao 2005). The dock protein plays a role in negatively regulating the insulin signaling pathway by facilitating the dephosphorylation of InR through recruitment of the ER-localized form of protein tyrosine phosphatase PTP61F, a function that is also observed in its mammalian ortholog *Nck* (Wu et al., 2011; Buszard et al., 2013)." (Bicanovsky et al., 2024).

"*D. ananassae* (NCBI:txid7217) is part of the *melanogaster* species group within the subgenus *Sophophora* of the genus *Drosophila* (Sturtevant 1939; Bock and Wheeler 1972). It was first described by Doeschall (1858). *D. ananassae* is circumtropical (Markow and O'Grady 2005; https://www.taxodros.uzh.ch, accessed 1 Feb 2023), and often associated with human settlement (Singh 2010). It has been extensively studied as a model for its cytogenetic and genetic characteristics, and in experimental evolution (Kikkawa 1938; Singh and Yadav 2015)." (Lawson et al., 2024).

We propose a gene model for the *D. ananassae* ortholog of the *D. melanogaster dreadlocks* (*dock*) gene. The genomic region of the ortholog corresponds to the uncharacterized protein XP_001961657.1 (LOC6497891) in the May 2011 (Agencourt dana_caf1/DanaCAF1) Genome Assembly of *D. ananassae* (GCA_000005115.1; Drosophila 12 Genomes Consortium et al., 2007). This model is based on RNA-Seq data from *D. ananassae* (SRP006203, SRP007906, PRJNA257286, PRJNA388952; Graveley et al., 2011) and *dock* in *D. melanogaster* using FlyBase release FB2023_03 (GCA_000001215.4; Larkin et al., 2021; Gramates et al., 2022; Jenkins et al., 2022).

Synteny

The reference gene, *dock*, occurs on chromosome 2L in *D. melanogaster* and is flanked upstream by *CG3662* and *CG3862* and downstream by *drongo* and *CG4291*. The *tblastn* search of *D. melanogaster* dock-PA (query) against the *D. ananassae* (GCA_000005115.1) Genome Assembly (database) placed the putative ortholog of *dock* within scaffold_12916 at locus LOC6497891 (XP_001961657.1)— with an E-value of 5e-171 and a percent identity of 83.54%. Furthermore, the putative ortholog is flanked upstream by LOC6497890 (XP_001961655.1) and LOC6497640 (XP_001961656.1), which correspond to *CG3662* and *CG3862* in *D. melanogaster* (E-value: 0.0 and 0.0; identity: 85.71% and 89.21%, respectively, as determined by *blastp*; Figure 1A; Altschul et al., 1990). The putative ortholog of *dock* is flanked downstream by LOC6497639 (XP_001961658.2) and LOC6497638 (XP_001961659.2), which correspond to *drongo* and *CG4291* in *D. melanogaster* (E-value: 0.0 and 0.0; identity: 74.27% and 84.66%, respectively, as determined by *blastp*). The putative ortholog assignment for *dock* in *D. ananassae* is supported by the following evidence: The genes surrounding the *dock* ortholog are orthologous to the genes at the same locus in *D. melanogaster* and local synteny is completely conserved, supported by results generated from *blastp*, so we conclude that LOC6497891 is the correct ortholog of *dock* in *D. ananassae* (Figure 1A).

Protein Model

dock in *D. ananassae* has four mRNA isoforms. *dock-RB*, *dock-RC*, and *dock-RD* are translated into identical proteins, and *dock-RA* is distinct in that it has a shorter first CDS (Figure 1B). All the RNA isoforms contain five CDSs in *D. melanogaster*. Relative to the ortholog in *D. melanogaster*, the protein isoform count is conserved (Figure 1B). There is some evidence for a novel isoform that is shorter than the others as predicted in the RefSeq genes track (XM_044716279), however, given the low support for the splice junction (20 reads) we default to the protocol's assumptions (Rele et al., 2023) that isoform structure is conserved relative to *D. melanogaster* unless there is strong evidence to the contrary, thus we are not annotating a novel isoform. The sequence of dock-PD in *D. ananassae* has 85.41% identity (E-value: 0.0) with the protein-coding isoform dock-PD in *D. melanogaster*, as determined by *blastp* (Figure 1C). Coordinates of this curated gene model (dock-PA, dock-PD, dock-PC, dock-PB) are stored by NCBI at GenBank (accessions BK064611, BK064612, BK064613, and BK064614). These data are also archived in the CaltechDATA repository (see “Extended Data” section below).

Methods

Detailed methods including algorithms, database versions, and citations for the complete annotation process can be found in Rele et al. (2023). Briefly, students use the GEP instance of the UCSC Genome Browser v.435 (https://gander.wustl.edu; Kent WJ et al., 2002; Navarro Gonzalez et al., 2021) to examine the genomic neighborhood of their reference IIS gene in the *D. melanogaster* genome assembly (Aug. 2014; BDGP Release 6 + ISO1 MT/dm6). Students then retrieve the protein sequence for the *D. melanogaster* reference gene for a given isoform and run it using *tblastn* against their target *Drosophila* species genome assembly on the NCBI BLAST server (https://blast.ncbi.nlm.nih.gov/Blast.cgi; Altschul et al., 1990) to identify potential orthologs. To validate the potential ortholog, students compare the local genomic neighborhood of their potential ortholog with the genomic neighborhood of their reference gene in *D. melanogaster*. This local synteny analysis includes at minimum the two upstream and downstream genes relative to their putative ortholog. They also explore other sets of genomic evidence using multiple alignment tracks in the Genome Browser, including BLAT alignments of RefSeq Genes, Spaln alignment of *D. melanogaster* proteins, multiple gene prediction tracks (e.g., GeMoMa, Geneid, Augustus), and modENCODE RNA-Seq from the target species. Detailed explanation of how these lines of genomic evidenced are leveraged by students in gene model development are described in Rele et al. (2023). Genomic structure information (e.g., CDSs, intron-exon number and boundaries, number of isoforms) for the *D. melanogaster* reference gene is retrieved through the Gene Record Finder (https://gander.wustl.edu/~wilson/dmelgenerecord/index.html; Rele et al., 2023). Approximate splice sites within the target gene are determined using *tblastn* using the CDSs from the *D. melanogaste*r reference gene. Coordinates of CDSs are then refined by examining aligned modENCODE RNA-Seq data, and by applying paradigms of molecular biology such as identifying canonical splice site sequences and ensuring the maintenance of an open reading frame across hypothesized splice sites. Students then confirm the biological validity of their target gene model using the Gene Model Checker (https://gander.wustl.edu/~wilson/genechecker/index.html; Rele et al., 2023), which compares the structure and translated sequence from their hypothesized target gene model against the *D. melanogaster* reference gene model. At least two independent models for a gene are generated by students under mentorship of their faculty course instructors. Those models are then reconciled by a third independent researcher mentored by the project leaders to produce the final model. Note: comparison of 5' and 3' UTR sequence information is not included in this GEP CURE protocol (Gruys et al., 2025).

Acknowledgements: This publication is dedicated to the memory of Dr. James J. Youngblom. We would like to thank Wilson Leung for developing and maintaining the technological infrastructure that was used to create this gene model. Also, thank you to Logan Cohen for assistance in updating the manuscript to the current template. Thank you to FlyBase for providing the definitive database for *Drosophila melanogaster* gene models.

Extended Data

Description: A GFF, FASTA, and PEP of the model. Resource Type: Model. File: DanaCAF1_dock.zip. DOI: 10.22002/5rcbb-mpv03

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215(3): 403-10. PubMed ID: 2231712

Bicanovsky GN, Lawson ME, Groeneveld A, Elkinton J, Dowell E, Shelley-Tremblay SF, et al., Reed LK. 2025. Gene model for the ortholog of dock in Drosophila eugracilis. microPublication Biology. DOI: 10.17912/micropub.biology.001848

Bock IR, Wheeler MR. 1972. The Drosophila melanogaster species group. Univ. Texas Publs Stud. Genet. 7(7213): 1-102. FBrf0024428

Buszard BJ, Johnson TK, Meng TC, Burke R, Warr CG, Tiganis T. 2013. The nucleus- and endoplasmic reticulum-targeted forms of protein tyrosine phosphatase 61F regulate Drosophila growth, life span, and fecundity. Mol Cell Biol 33(7): 1345-56. PubMed ID: 23339871

Doleschall CL. 1858. Derde bijdrage tot de kennis der Dipteren fauna van nederlandsch indie. Natuurk. Tijd. Ned.-Indie 17: 73-128. FBrf0000091

Drosophila 12 Genomes Consortium, Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, et al., MacCallum I. 2007. Evolution of genes and genomes on the Drosophila phylogeny. Nature 450(7167): 203-18. PubMed ID: 17994087

Garrity PA, Rao Y, Salecker I, McGlade J, Pawson T, Zipursky SL. 1996. Drosophila photoreceptor axon guidance and targeting requires the dreadlocks SH2/SH3 adapter protein. Cell 85(5): 639-50. PubMed ID: 8646773

Gramates LS, Agapite J, Attrill H, Calvi BR, Crosby M, dos Santos G Goodman JL, Goutte-Gattat D, Jenkins V, Kaufman T, Larkin A, Matthews B, Millburn G, Strelets VB, and the FlyBase Consortium (2022) FlyBase: a guided tour of highlighted features. Genetics, Volume 220, Issue 4, April 2022. DOI: 10.1093/genetics/iyac035

Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, et al., Celniker SE. 2011. The developmental transcriptome of Drosophila melanogaster. Nature 471(7339): 473-9. PubMed ID: 21179090

Grewal SS. 2009. Insulin/TOR signaling in growth and homeostasis: a view from the fly world. Int J Biochem Cell Biol 41(5): 1006-10. PubMed ID: 18992839

Gruys ML, Sharp MA, Lill Z, Xiong C, Hark AT, Youngblom JJ, Rele CP, Reed LK. 2025. Gene model for the ortholog of Glys in Drosophila simulans. MicroPubl Biol 2025: 10.17912/micropub.biology.001168. PubMed ID: 39845267

Hietakangas V, Cohen SM. 2009. Regulation of tissue growth through nutrient sensing. Annu Rev Genet 43: 389-410. PubMed ID: 19694515

Hing H, Xiao J, Harden N, Lim L, Zipursky SL. 1999. Pak functions downstream of Dock to regulate photoreceptor axon guidance in Drosophila. Cell 97(7): 853-63. PubMed ID: 10399914

Jenkins VK, Larkin A, Thurmond J, FlyBase Consortium. 2022. Using FlyBase: A Database of Drosophila Genes and Genetics. Methods Mol Biol 2540: 1-34. PubMed ID: 35980571

Kaipa BR, Shao H, Schäfer G, Trinkewitz T, Groth V, Liu J, et al., Önel SF. 2013. Dock mediates Scar- and WASp-dependent actin polymerization through interaction with cell adhesion molecules in founder cells and fusion-competent myoblasts. J Cell Sci 126(Pt 1): 360-72. PubMed ID: 22992459

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. Genome Res 12(6): 996-1006. PubMed ID: 12045153

Kikkawa H. 1938. Studies on the genetics and cytology ofDrosophila ananassae. Genetica 20: 458-516. DOI: 10.1007/BF01531779

Larkin A, Marygold SJ, Antonazzo G, Attrill H, Dos Santos G, Garapati PV, et al., FlyBase Consortium. 2021. FlyBase: updates to the Drosophila melanogaster knowledge base. Nucleic Acids Res 49(D1): D899-D907. PubMed ID: 33219682

Lawson ME, McAbee M, Lucas RA, Tanner S, Wittke-Thompson J, Pelletier TA, et al., Reed LK. 2024. Gene model for the ortholog of Ilp5 in Drosophila ananassae. MicroPubl Biol 2024: 10.17912/micropub.biology.000782. PubMed ID: 39717145

Markow TA, O'Grady P. 2005. *Drosophila*: A guide to species identification and use. Academic Press 978-0-12-473052-6

Mudge JM, Harrow J. 2016. The state of play in higher eukaryote gene annotation. Nat Rev Genet 17(12): 758-772. PubMed ID: 27773922

Myers A, Hoffman A, Natysin M, Arsham AM, Stamm J, Thompson JS, Rele CP, Reed LK. 2024. Gene model for the ortholog Myc in Drosophila ananassae. MicroPubl Biol 2024: 10.17912/micropub.biology.000856. PubMed ID: 39677519

Navarro Gonzalez J, Zweig AS, Speir ML, Schmelter D, Rosenbloom KR, Raney BJ, et al., Kent WJ. 2021. The UCSC Genome Browser database: 2021 update. Nucleic Acids Res 49(D1): D1046-D1057. PubMed ID: 33221922

Raney BJ, Dreszer TR, Barber GP, Clawson H, Fujita PA, Wang T, et al., Kent WJ. 2014. Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. Bioinformatics 30(7): 1003-5. PubMed ID: 24227676

Rao Y, Zipursky SL. 1998. Domain requirements for the Dock adapter protein in growth- cone signaling. Proc Natl Acad Sci U S A 95(5): 2077-82. PubMed ID: 9482841

Rao Y. 2005. Dissecting Nck/Dock signaling pathways in Drosophila visual system. Int J Biol Sci 1(2): 80-6. PubMed ID: 15951852

Rele CP, Sandlin KM, Leung W, Reed LK. 2022. Manual annotation of Drosophila genes: a Genomics Education Partnership protocol. F1000Res 11: 1579. PubMed ID: 37854289

Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, Dixon JE, Zipursky SL. 2000. Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell 101(6): 671-84. PubMed ID: 10892653

Singh BN, Yadav JP. 2015. Status of research on Drosophila ananassae at global level. J Genet 94(4): 785-92. PubMed ID: 26690536

Singh BN. 2010. Drosophila ananassae: a good model species for genetical, behavioural and evolutionary studies. Indian J Exp Biol 48(4): 333-45. PubMed ID: 20726331

Song J, Wu L, Chen Z, Kohanski RA, Pick L. 2003. Axons guided by insulin receptor in Drosophila visual system. Science 300(5618): 502-5. PubMed ID: 12702880

Stark K, Crowe O, Lewellyn L. 2021. Precise levels of the Drosophila adaptor protein Dreadlocks maintain the size and stability of germline ring canals. J Cell Sci 134(8). PubMed ID: 33912915

Stevens A, Jacobs JR. 2002. Integrins regulate responsiveness to slit repellent signals. J Neurosci 22(11): 4448-55. PubMed ID: 12040052

Sturtevant AH. 1939. On the Subdivision of the Genus Drosophila. Proc Natl Acad Sci U S A 25(3): 137-41. PubMed ID: 16577879

Tello-Ruiz MK, Marco CF, Hsu FM, Khangura RS, Qiao P, Sapkota S, et al., Micklos DA. 2019. Double triage to identify poorly annotated genes in maize: The missing link in community curation. PLoS One 14(10): e0224086. PubMed ID: 31658277

Weng YL, Liu N, DiAntonio A, Broihier HT. 2011. The cytoplasmic adaptor protein Caskin mediates Lar signal transduction during Drosophila motor axon guidance. J Neurosci 31(12): 4421-33. PubMed ID: 21430143

Willoughby LF, Manent J, Allan K, Lee H, Portela M, Wiede F, et al., Richardson HE. 2017. Differential regulation of protein tyrosine kinase signalling by Dock and the PTP61F variants. FEBS J 284(14): 2231-2250. PubMed ID: 28544778

Wu CL, Buszard B, Teng CH, Chen WL, Warr CG, Tiganis T, Meng TC. 2011. Dock/Nck facilitates PTP61F/PTP1B regulation of insulin signalling. Biochem J 439(1): 151-9. PubMed ID: 21707536

Funding: This material is based upon work supported by the National Science Foundation under Grant No. IUSE-1915544 to LKR and the National Institute of General Medical Sciences of the National Institutes of Health Award R25GM130517 to LKR. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
Supported by National Institutes of Health (United States) R25GM130517 to LK Reed.
Supported by National Science Foundation (United States) 1915544 to LK Reed.

Author Contributions: Anne E. Backlund: formal analysis, validation, writing - original draft, writing - review editing. Tamica D'Souza: formal analysis, writing - review editing. Jaskamaldip Kaur: formal analysis, writing - review editing. Anya Goodman: supervision, writing - review editing. James J. Youngblom: supervision, writing - review editing. Chinmay P. Rele: data curation, formal analysis, methodology, project administration, software, supervision, validation, visualization, writing - review editing. Laura K Reed: supervision, funding acquisition, conceptualization, project administration, writing - review editing, methodology.

Reviewed By: Anonymous

History: Received October 11, 2023 **Revision Received** January 7, 2025 **Accepted** September 12, 2025 **Published Online** September 14, 2025 **Indexed** September 28, 2025

Citation: Backlund AE, D'Souza T, Kaur J, Goodman A, Youngblom JJ, Rele CP, Reed LK. 2025. Gene model for the ortholog of *dock* in *Drosophila ananassae*. microPublication Biology. 10.17912/micropub.biology.001025